

Intra-chromosomal estimates of inbreeding and coancestry in the Spanish Holstein cattle population

D. Kleinman-Ruiz, B. Villanueva, J. Fernández, M.A. Toro, L.A. García-Cortés, S.T. Rodríguez-Ramilo

ABSTRACT

In recent years, inbreeding and coancestry are being estimated from genome-wide molecular information using a large number of SNPs. Molecular inbreeding and coancestry can be calculated for the whole genome or for particular regions of the genome. In this study, genome-based inbreeding and coancestry were estimated per chromosome and at intra-chromosomal level in a group of Holstein animals genotyped with the Illumina BovineSNP50 BeadChip. After applying filtering criteria, the genomic dataset included 36,693 autosomal SNPs and 10,569 animals. Genome-based inbreeding and coancestry at intra-chromosomal level were calculated using sliding windows of approximately 5 Mb. The results showed differential patterns of inbreeding and coancestry on specific chromosome regions. These patterns provide a more detailed picture of genetic diversity that could be used, for example, for the detection of regions with low levels of genetic diversity that require a specific genetic management in conservation programmes.

Keywords:
Genome-wide information
Holstein
Coancestry
Inbreeding
Genomic regions

1. Introduction

Over the last decades, inbreeding and coancestry in livestock populations have been alternatively measured using pedigree information or microsatellite data when pedigree information was not reliable or available. Currently, the availability of high-density SNP chips allows a more detailed evaluation of inbreeding and coancestry based on genomic estimates. The reasons why these measures are expected to be more accurate than the pedigree-based ones are basically as follows. First, genomic estimates reflect the percentage of homozygous positions given the genomic information, or the realised pairwise coancestry, while pedigree based estimates are just the expectations of these measures. Second, genomic estimates can detect relationships due to very distant common ancestors, which pedigree-based estimates do not take into account (Keller et al., 2011). An additional advantage of genome-based estimates is that they can be calculated for particular regions of the genome.

Published studies have shown that the variation in genetic diversity across regions could be substantially large (e.g. The International SNP Map Working Group, 2001; Engelsma et al., 2012; Esteve-Codina et al., 2013). The causes for these patterns are diverse. For example, it is well known that changes in genotype

frequencies caused by selection affect the frequencies of neutral variants and other genetically linked sites in the genome, according to the theory of 'hitch-hiking' (Maynard-Smith and Haigh, 1974). As the favourable (unfavourable) allele in the selected locus increases (decreases) its frequency in a population, in the surrounding loci a parallel increase in frequency of one allele can also be observed due to the linkage between the selected locus and the rest. Consequently, there will be a greater loss of genetic diversity in regions harbouring selective loci and higher levels of inbreeding and coancestry.

In recent times, due to the rapid development of molecular resources, large numbers of SNPs have been used for the implementation of genome-wide evaluations (Meuwissen et al., 2001) in many commercial breeding programmes, including Holstein populations (VanRaden and Cooper, 2011). Despite the higher accuracy of genome-wide evaluations (which could lead to more intense selection and a faster increase in inbreeding), Daetwyler et al. (2007) showed that the global levels of diversity maintained per generation are higher when performing genome-wide selection than when using conventional BLUP based on pedigree relationships. However, Heidari et al. (2014) showed that genome-wide selection applies selection pressure much more locally than BLUP, resulting in larger allele frequency changes in the selected loci and the linked ones. Therefore, the loss of genetic diversity on specific genomic regions associated to selection processes may be larger when genome-wide selection is implemented.

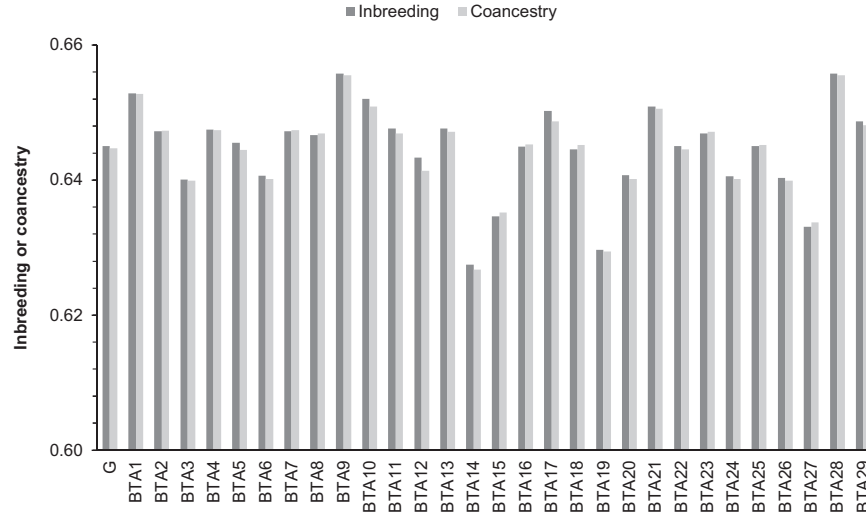


Fig. 1. Averaged inbreeding and coancestry across the genome (G) and for each autosome (BTA1-BTA29).

On the other hand, genomic regions harbouring loci involved in general disease resistance (such the MHC, where a high level of genetic diversity ensures that the population can deal with potential new disease challenges) may show low levels of inbreeding and coancestry (Birch et al., 2006). Both types of genomic regions could be essential for the potential adaptation and survival of any population.

Moreover, even the action of genetic drift may lead to some genomic regions having less genetic variation than others. Therefore, it may be arguable that regions with lower levels of diversity should be detected and specific management implemented to control diversity on such regions (Gómez-Romano et al., 2014). Genome-wide inbreeding and coancestry could be useful to detect such variation across the genome.

The objective of this study was to evaluate the variation of inbreeding and coancestry based on SNP marker information over the whole genome, at the chromosomal level and within chromosomes. For this purpose we used data from individuals belonging to the Spanish Holstein population and genotyped for the Illumina BovineSNP50 BeadChip.

2. Materials and methods

2.1. Genomic data

Genomic data used in this study were the same as in Rodríguez-Ramilo et al. (2015). Genomic information from 11,135 animals belonging to the Spanish Holstein population was analysed. These individuals were genotyped with the Illumina BovineSNP50 BeadChip (versions v1 or v2). Only SNPs common to both chip versions were selected for the analysis (52,340 SNPs). SNP positions within the genome were obtained from the UMD 3.0 bovine genome assembly. Unmapped SNPs (523) and those mapped on chromosomes X or Y (1056) were excluded. In addition, 14,068 SNPs with missing genotypes for more than 5% of the individuals were discarded. After that, 566 animals with more than 5% missing genotypes for the remaining 36,693 SNPs were also removed. The final dataset included 36,693 autosomal SNPs and 10,569 animals (9990 bulls and 579 cows).

2.2. Estimates of coancestry and inbreeding coefficients

The molecular coancestry coefficient between individuals i and j (f_{ij}) at a given locus can be defined as the probability that two

alleles taken at random from each individual are alike in state. When dealing with several loci, it is the average across loci. In this study, f_{ij} was calculated following Nejati-Javaremi et al. (1997) as

$$f_{ij} = (1/n_s) \sum_{s=1}^{n_s} \left[\left(\sum_{k=1}^2 \sum_{m=1}^2 I_{s_{kijm}} \right) / 4 \right],$$

where n_s is the number of SNPs used and $I_{s_{kijm}}$ is the identity of the k th allele from individual i with the m th allele from individual j at SNP s , and takes a value of 1 if both alleles are identical and 0 otherwise. For a single locus, the molecular inbreeding coefficient of individual i (F_i) is 1 if the individual is homozygous for this locus and 0 if it is heterozygous. For a group of loci, F_i is the proportion of homozygous loci. The inbreeding coefficient of individual i was calculated in this study as $F_i = 2f_{ii} - 1$.

Chromosomal estimates of inbreeding and coancestry were obtained using all the SNPs genotyped on each chromosome after applying filtering criteria. The number of genotyped SNPs was different between chromosomes, however the density of SNPs was very similar across chromosomes (data not shown).

Within chromosomes, estimates of inbreeding and coancestry were also obtained for different regions. Instead of calculating coefficients for individual SNPs, inbreeding and coancestry were estimated over sliding windows with a window size of approximately 5 Mb. Following this approach, the noise of single-locus estimates can be reduced by combining data from several adjacent markers. This procedure was based on the method used by Weir et al. (2005) and Engelsma et al. (2012). For each chromosome, the first window was identified by taking the SNPs at the first 5 Mb of the chromosome. Subsequently, the window moves along the genetic map in single-SNP increments, until the end of the chromosome is reached. In each window, the same number of SNPs for that specific chromosome has been maintained. The average number of SNPs in each window across chromosomes was 70.76 ± 3.00 . It must be pointed out that, in that way, window size will not always be exactly 5 Mb. For each window, inbreeding and coancestry were estimated by averaging the values for all SNPs lying in that window. Afterwards, values were averaged over animals (or pairs of animals). A total of $(10,569 \times 10,568)/2$ pairs for coancestry estimates.

To detect regions showing significantly different levels of inbreeding and coancestry, their frequency distribution was plotted for each chromosome. As observed by Weir et al. (2005) for F_{ST} values across the genome, if it is assumed that differences in coancestry

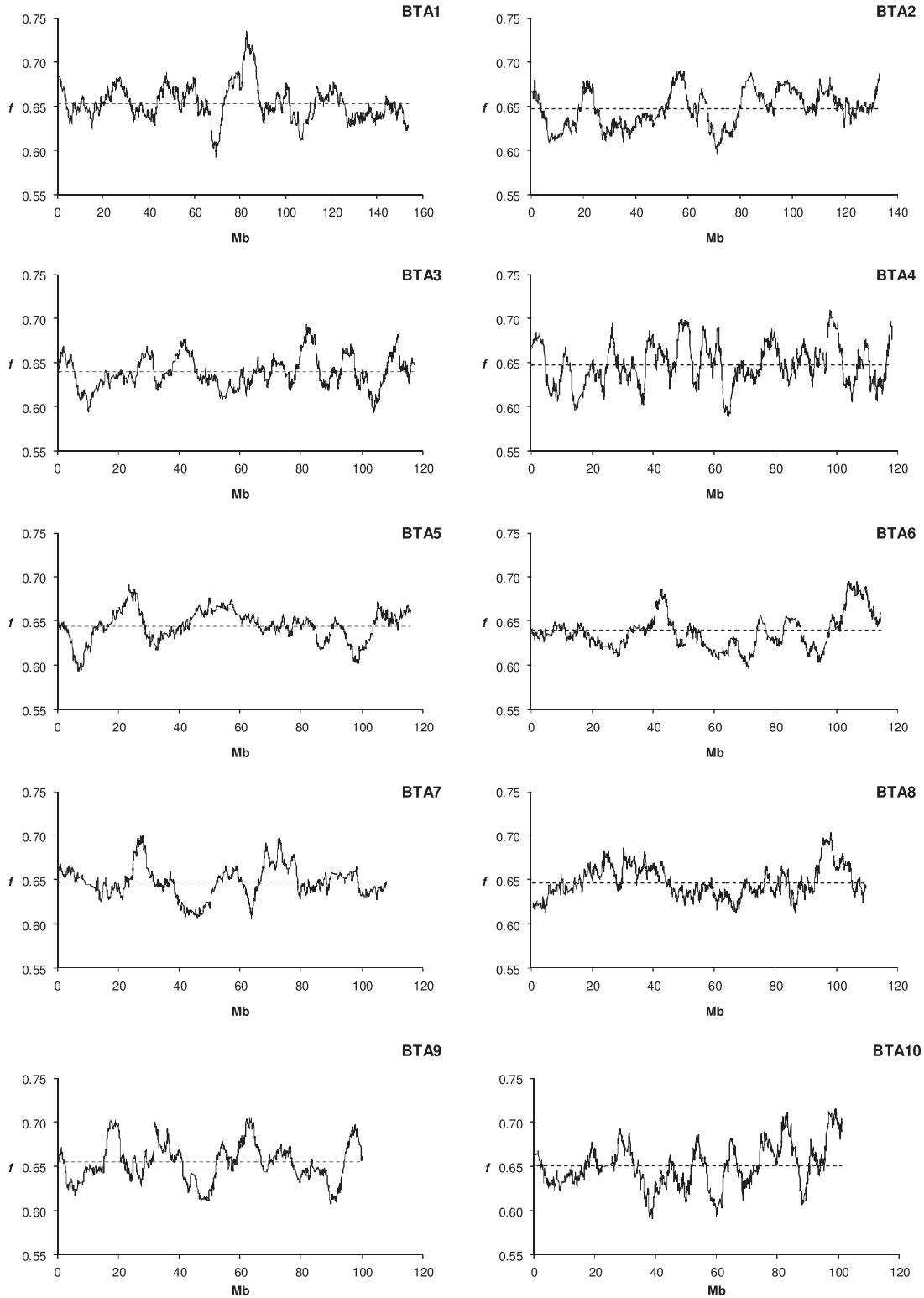


Fig. 2. Coancestry patterns along the 29 *Bos taurus* autosomes (BTA1-BTA29). The dashed line indicates the averaged coancestry for each autosome.

were only due to genetic drift (sampling), distributions should be normal. Departures from normality will point out the existence of regions with inbreeding or coancestry significantly higher or lower than the chromosome average. Asymmetric or leptokurtic distributions (weighty tails) may be due to such differential regions. To

determine the levels of significance, Kolmogorov–Smirnov normality tests were performed for each chromosome. It must be highlighted that inbreeding and coancestry at adjacent windows are not independent due to the linkage disequilibrium and, thus, corrections for auto-correlated data have been implemented.

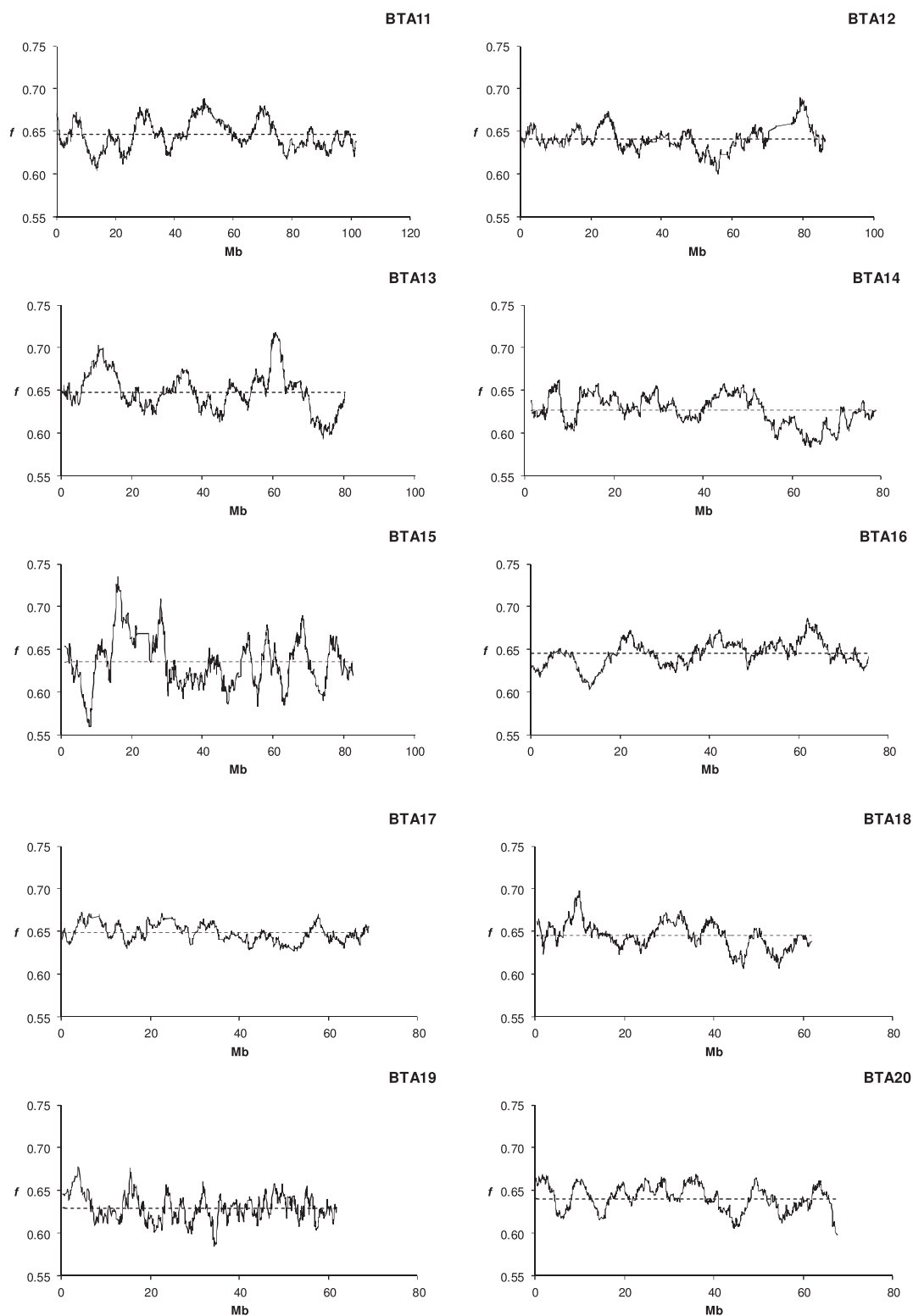


Fig. 2. (continued)

3. Results

Mean genome-based estimates of inbreeding and coancestry were 0.6451 ($\sigma^2=0.0002$) and 0.6447 ($\sigma^2=0.0001$), respectively, for the whole genome.

3.1. Chromosomal estimates

Fig. 1 shows the genomic inbreeding and coancestry across chromosomes. In general, inbreeding and coancestry values were very similar at the chromosome level. The lowest value was

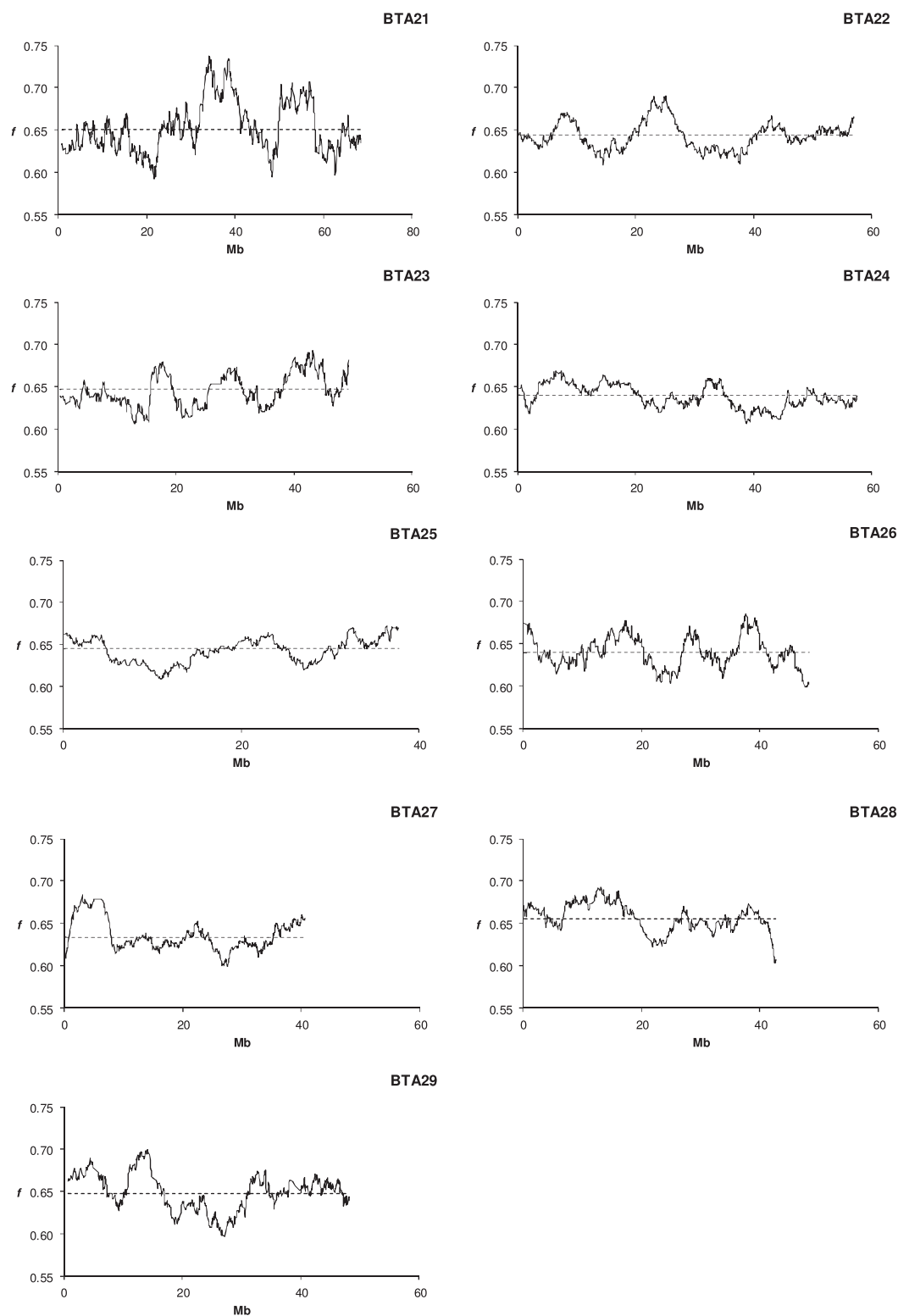


Fig. 2. (continued)

observed in chromosome 14 ($F \sim f = 0.627$), and the highest values were observed for chromosomes 9 and 28 ($F \sim f = 0.656$).

3.2. Intra-chromosomal genomic patterns

Similar patterns were observed for inbreeding and coancestry and, thus, only coancestry results are presented. When coancestry coefficients were smoothed by using a sliding window approach,

differences in coancestry within chromosomes were clear at specific chromosome regions (Fig. 2). Table 1 shows the mean, maximum and minimum coancestry values for each autosome. Autosome 15 showed the highest range between coancestry values and chromosome 17 showed the lowest difference between the minimum and the maximum coancestry values.

Fig. 3a shows in more detail the molecular coancestry along the different windows in chromosome BTA1. Coancestry values across

Table 1
Mean, minimum and maximum coancestry values for each autosome with the sliding window approach.

BTA	Mean	Minimum	Maximum
1	0.653	0.593	0.735
2	0.646	0.595	0.691
3	0.640	0.593	0.694
4	0.648	0.590	0.711
5	0.644	0.594	0.691
6	0.639	0.596	0.695
7	0.648	0.605	0.701
8	0.648	0.612	0.704
9	0.656	0.608	0.705
10	0.650	0.592	0.716
11	0.646	0.604	0.689
12	0.642	0.600	0.689
13	0.647	0.594	0.717
14	0.626	0.584	0.662
15	0.635	0.560	0.736
16	0.645	0.604	0.686
17	0.648	0.627	0.672
18	0.645	0.607	0.697
19	0.630	0.585	0.677
20	0.642	0.598	0.669
21	0.652	0.592	0.737
22	0.643	0.609	0.690
23	0.646	0.607	0.692
24	0.638	0.607	0.669
25	0.643	0.609	0.673
26	0.640	0.599	0.685
27	0.633	0.599	0.684
28	0.658	0.603	0.692
29	0.648	0.597	0.700

For each chromosome, the sliding windows comprise a fixed number of SNPs. The first window was identified by taking the SNPs at the first 5 Mb of the chromosome. Subsequent windows move along the genetic map in single-SNP increments until the end of the chromosome is reached.

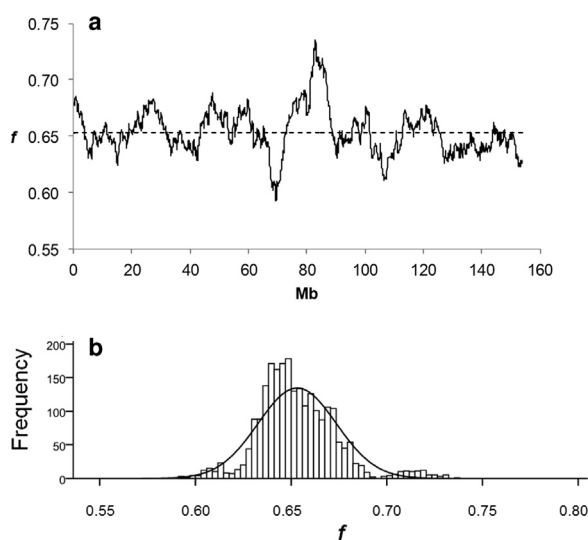


Fig. 3. Coancestry patterns (a) and frequency distribution (b) on chromosome BTA1. The dashed line indicates the averaged coancestry.

regions varied substantially within the chromosome, presenting regions with higher and lower coancestries than the average for that chromosome. This was also the pattern for the rest of chromosomes, although in some cases differences between regions were smaller (Fig. 2).

Fig. 3b presents the frequencies distribution of coancestry for BTA1. This graph shows a nearly bimodal pattern, with a ‘heavy’ upper tail. This could suggest the existence of a region in that chromosome having a significantly higher coancestry than the rest of the chromosome. However, normality test did not reach statistical significance ($p > 0.01$). Frequencies distributions for all the autosomes are shown in Fig. 4. None of the Kolmogorov–Smirnov tests reached significance.

4. Discussion

In this study, molecular inbreeding and coancestry coefficients were calculated for the whole genome as well as per chromosome and within chromosomes using genomic data from bulls and cows belonging to the Spanish Holstein population.

The results indicate that genome-based measures of inbreeding and coancestry are able to detect differences between chromosome regions, providing a more detailed picture of the genetic diversity. In this situation, it could be possible to focus on the specific control of inbreeding and coancestry on certain genomic regions (Gómez-Romano et al., 2014) because diversity at those regions is already low (due to previous selection processes) or because it harbours genes for which populations with greater levels of diversity exhibit higher fitness (e.g. MHC genes). Gómez-Romano et al. (2014) showed, through computer simulations, that the optimal contributions method is very efficient in maintaining diversity (measured as expected heterozygosity) in specific regions of the genome when using a coancestry coefficient computed using the SNPs mapped on those regions. The method also allows to efficiently restrict the loss of diversity across the rest of the genome. Therefore, the management of diversity in conservation programmes as well as in genomic selection programmes could be improved with the inclusion of this chromosomal and intra-chromosomal information about inbreeding and coancestry.

It is expected that neutral loci close to quantitative trait loci subjected to selection have higher levels of inbreeding and coancestry than other further apart. Some studies carried out in cattle have compared breeds with different production aptitudes, for example dairy and beef breeds (e.g. Hayes et al., 2008). These studies highlighted genes with an effect on coancestry and phenotypes (e.g. *DGAT1* on autosome 14 for dairy and *MSTN* on autosome 2 for beef cattle, respectively). Several studies have also compared breeds with similar production aptitudes. For example, Engelsma et al. (2012) found several regions with substantial differences in diversity between a high and a low production Holstein groups.

The differences of coancestry patterns in chromosomal regions detected in this study could highlight the particular effect of artificial selection in those genomic regions with differences in genetic diversity. However, statistical significance for normality tests with auto-correlated data has not been reached. Further research should be carried out in order to elucidate whether the genome-wide measure of coancestry and inbreeding could be used to detect selection signatures. The inbreeding coefficient based on runs of homozygosity (ROH) has been suggested to be useful for detecting selective sweeps (Hillestad, 2015). However, there are still doubts about the way of defining those ROH and the consequences of different criteria on the efficiency of detection.

Conflicts of interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have

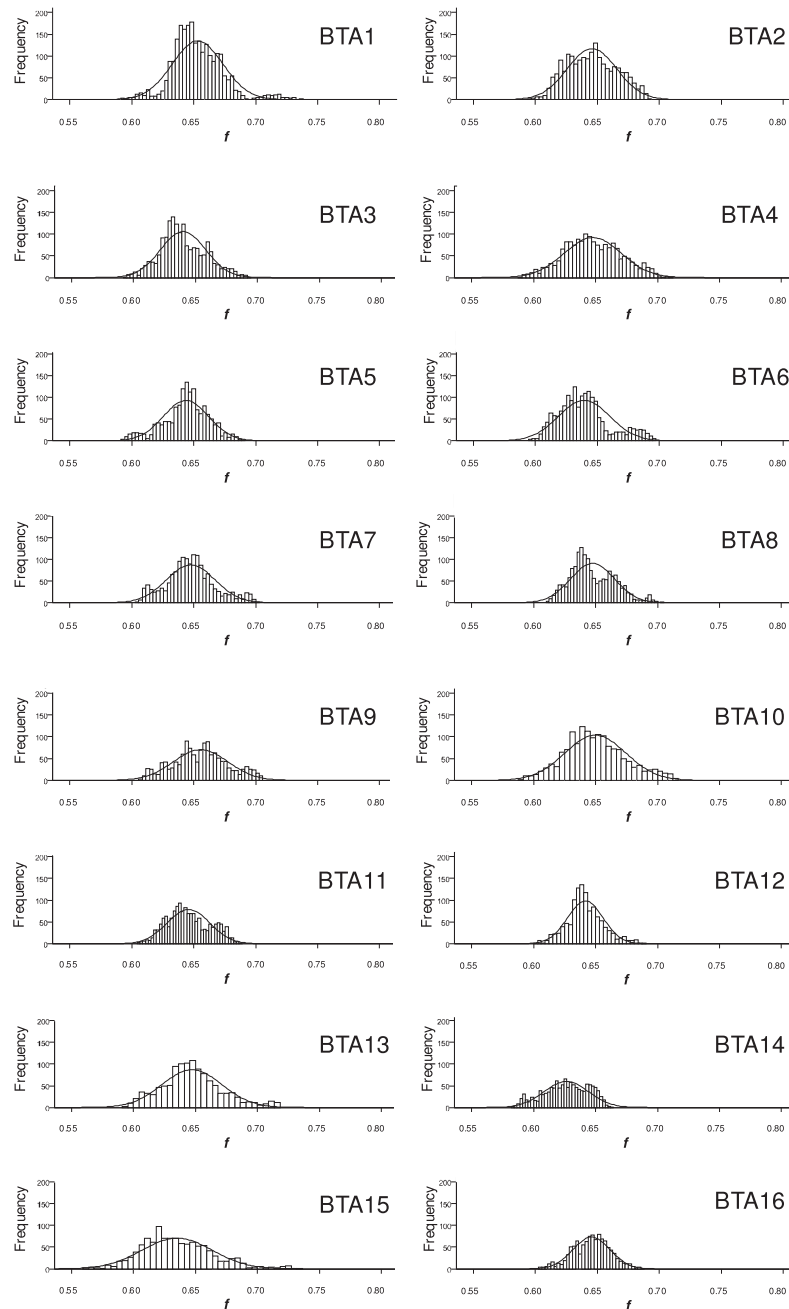


Fig. 4. Frequencies distribution of coancestry on *Bos taurus* autosomes (BTA1-BTA29).

influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and

direct communications with the office). She is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from the editorial team of Livestock Science.

Acknowledgements

We are very grateful to the EuroGenomics Consortium and Conafe for providing genotypes used in this study. This study was supported by the Ministerio de Economía y Competitividad, Spain (Grant CGL2012-39861-C02-02).

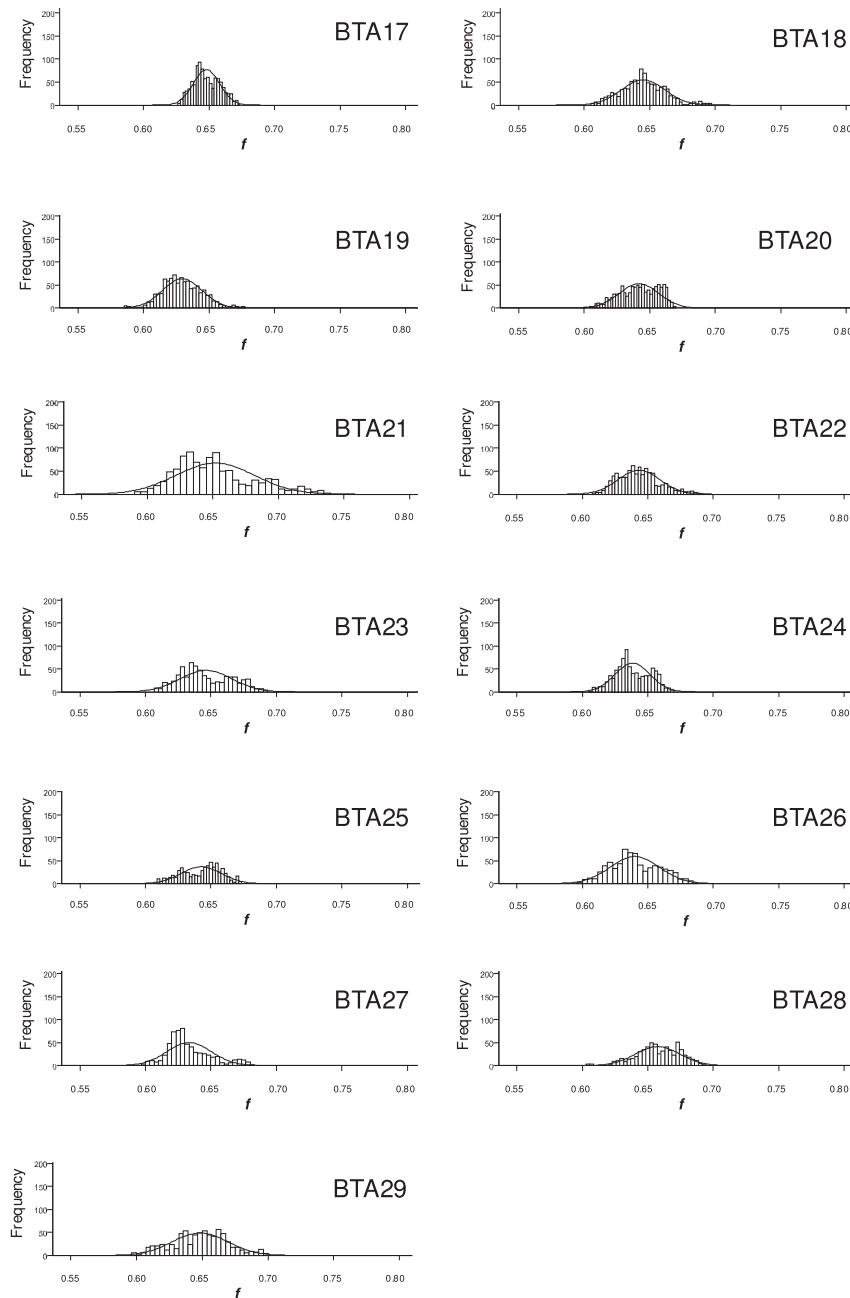


Fig. 4. (continued)

References

- Birch, J., Murphy, L., MacHugh, N.D., Ellis, S.A., 2006. Generation and maintenance of diversity in the cattle MHC class I region. *Immunogenetics* 58, 670–679.
- Daetwyler, H.D., Villanueva, B., Bijma, P., Woolliams, J.A., 2007. Inbreeding in genome-wide selection. *J. Anim. Breed. Genet.* 124, 369–376.
- Engelsma, K.A., Veerkamp, R.F., Calus, M.P.L., Bijma, P., Windig, J.J., 2012. Pedigree and marker-based methods in the estimation of genetic diversity in small groups of Holstein cattle. *J. Anim. Breed. Genet.* 129, 195–205.
- Esteve-Codina, A., Paudel, Y., Ferretti, L., Raineri, E., Megens, H.-J., Silió, L., Rodríguez, C., Groenen, M.A.M., Ramos-Onsins, S.E., Pérez-Enciso, M., 2013. Dissecting structural and nucleotide genome-wide variation in inbred Iberian pigs. *BMC Genomics* 14, 148.
- Gómez-Romano, F., Villanueva, B., Fernández, J., Woolliams, J.A., Pong-Wong, R., 2014. The use of genomic coancestry matrices in the optimisation of contributions for maintaining genetic diversity at specific regions of the genome. In: *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production*. Vancouver, Canada.
- Hayes, B.J., Chamberlain, A.J., Maceachern, S., Savin, K., McPartlan, H., MacLeod, I., 2008. A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. *Anim. Genet.* 40, 176–184.
- Heidaritabar, M., Vereijken, A., Muir, W.M., Meuwissen, T.H.E., Cheng, H., Megens, H.J., Groenen, M.A.M., Bastiaansen, J.W.M., 2014. Systematic differences in the response of genetic variation to pedigree and genome-based selection methods. *Heredity* 113, 503–513.
- Hillestad, B., 2015. Inbreeding Determined by the Amount of Homozygous Regions in the Genome. Norwegian University of Life Sciences, Ås, Norway.
- Keller, M.C., Visscher, P.M., Goddard, M.E., 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189, 237–249.
- Maynard-Smith, J., Haigh, J., 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* 23, 23–35.
- Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value

using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Nejati-Javaremi, A., Smith, C., Gibson, J.P., 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. *J. Anim. Sci.* 75, 1738–1745.

Rodríguez-Ramilo, S.T., Fernández, J., Toro, M.A., Hernández, D., Villanueva, B., 2015. Genome-wide estimates of coancestry, inbreeding and effective population size in the Spanish Holstein population. *Plos ONE* 10, e0124157.

The International SNP Map Working Group, 2001. A map of human genome

sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928–933.

VanRaden, P.M., Cooper, T., 2011. The genomic evaluation system in the United States: past, present, future. *J. Dairy Sci.* 94, 3202–3211.

Weir, B.S., Cardon, L.R., Anderson, A.D., Nielsen, D.M., Hill, W.G., 2005. Measures of human population structure show heterogeneity among genomic regions. *Genome Res.* 15, 1468–1476.